BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

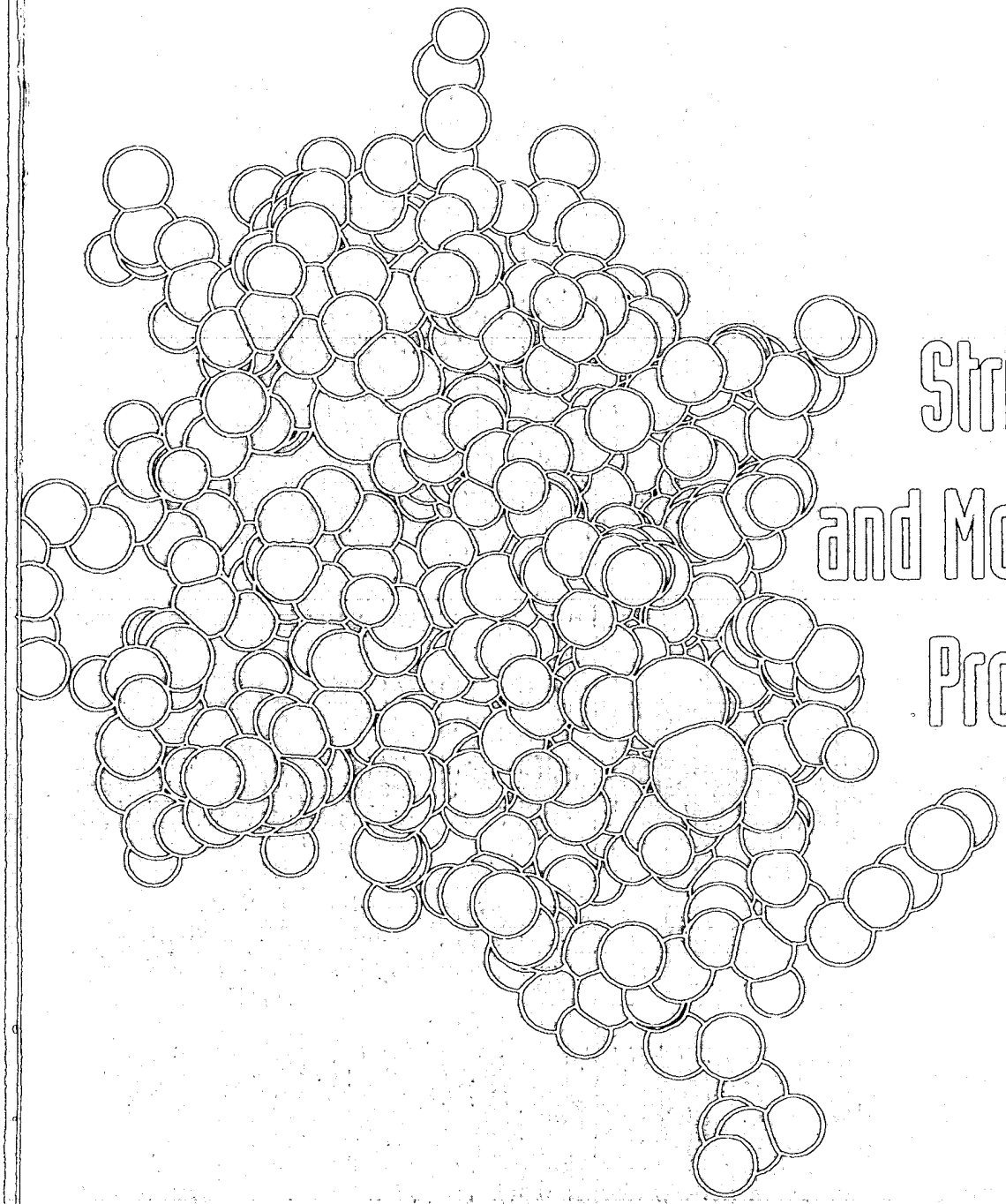
- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

SECOND EDITION

PROTEINS



Structures and Molecular Properties

THOMAS E. CREIGHTON

Cover image provided by Martin Noble.

Library of Congress Cataloging-in-Publication Data
Creighton, Thomas E., 1940–
Proteins : structures and molecular properties / Thomas E. Creighton. — 2nd ed.
p. cm.
Includes bibliographical references and index.
ISBN 0-7167-7030-X
1. Proteins—Structure. 2. Proteins—Chemistry. I. Title.
QP551.C737 1993
574.19'245—dc20 92-6664
CIP

Printed in the United States of America

3 4 5 6 7 8 9 0 KP 9 9 8 7 6 5 4

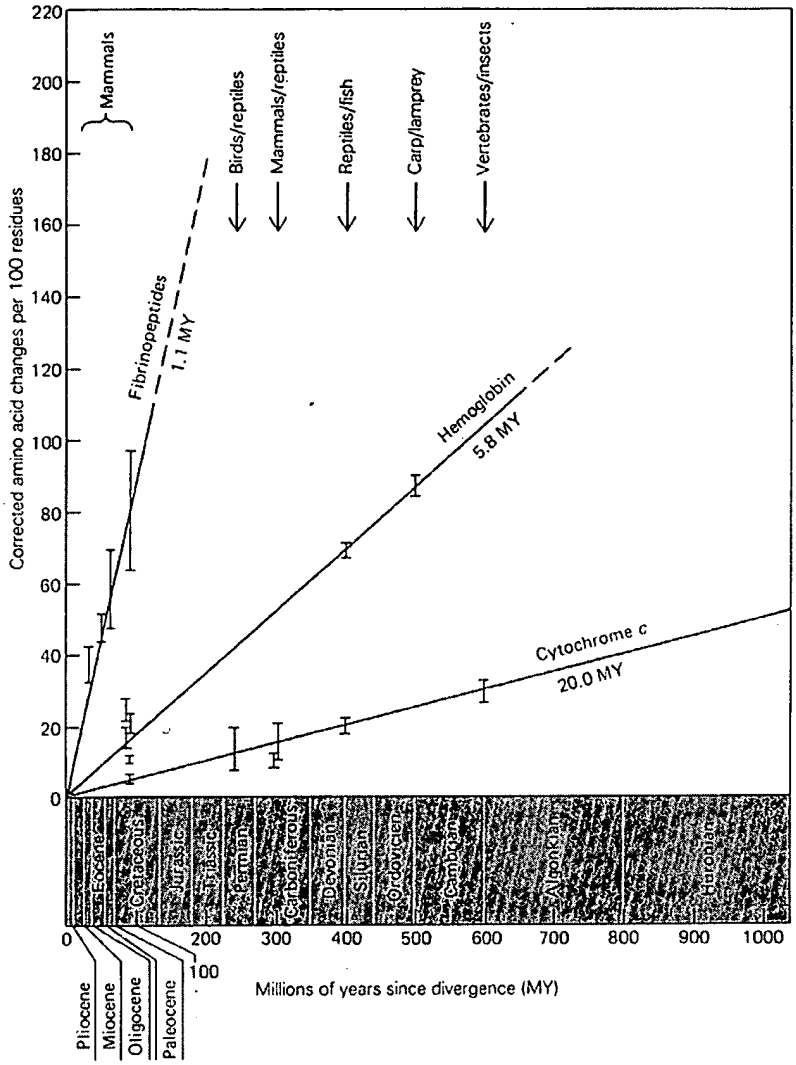


FIGURE 3.11
Rates of evolutionary divergence of fibrinopeptides A and B, hemoglobin-α and -β chains, and cytochromes *c*. The number of amino acid changes per 100 residues, corrected for multiple changes, is plotted versus the estimated time since separation of the genes for the proteins compared. Below each line is the *unit evolutionary period* in millions of years (MY), the time required for a change of 1% of the residues. The times of divergence of some evolutionary lineages are indicated at the top. (Adapted from R. E. Dickerson, *J. Mol. Evol.* 1:26–45, 1971.)

Evidence for higher rates of nucleotide substitution in rodents than in man. C. I. Wu and W. H. Li. *Proc. Natl. Acad. Sci. USA* 82:1741–1745 (1985).

Evolution of cytochrome *c* genes and pseudogenes. C. I. Wu et al. *J. Mol. Evol.* 23:61–75 (1986).

Molecular time scale for evolution. A. C. Wilson et al. *Trends Genet.* 3:241–247 (1987).

Rate constancy of globin gene evolution in placental mammals. S. Easteal. *Proc. Natl. Acad. Sci. USA* 85:7622–7626 (1988).

d. Roles of Selection

If mutations occur at a constant rate in all genes, how can we explain the wide range of evolutionary rates of change among different proteins (Table 3.3) and the nearly constant rate for each protein? The most plausible explanation is that the observed differences among proteins are largely due to neutral mutations that do not significantly affect protein function and so have not been selected for or against. This is not to say

that natural selection has not been important, because it certainly must have selected against adverse mutations.

According to the neutral mutation hypothesis, the constant rate of divergence of a protein is the same as its particular neutral mutation rate per gene copy, which is the total mutation rate times the fraction of mutations that are effectively neutral. Even if the total mutation rate is the same for all genes, the neutral mutation rate would differ for each gene because of the different fractions of mutations in the various proteins that are effectively neutral; every gene and protein would differ from every other in how much its amino acid sequence can vary without affecting its function. If the exact amino acid sequence is not critical for the function of a protein, a large fraction of its total mutations would be neutral, and the sequence of the protein would evolve rapidly. Fibrinopeptides are examples of such proteins. They appear to function primarily to block the aggregation of the precursor protein, fibrinogen. They are cleaved proteolytically from the amino ends of two of the three fibrinogen polypeptides in the first step of blood clotting and play no further known role. As a consequence of their removal, the fibrinogen is converted to fibrin, which aggregates and forms the framework of the blood clot. The only known functional constraints on the amino acid sequences of the fibrinopeptides are a carboxyl-terminal Arg residue, which is required for proteolytic cleavage by thrombin, and a somewhat acidic net charge, which probably inhibits aggregation of the precursor, fibrinogen. Within these minor limitations, many amino acid sequences are functional, which explains why these protein segments have evolved at relatively rapid rates (Table 3.3).

At the other extreme, proteins for which very few amino acid replacements are acceptable evolve at very slow rates. An example is cytochrome *c*, which must interact with a number of other proteins in its function of transferring electrons (Sec. 8.3.4.b, Table 3.3). Variation has occurred at only a few sites (Fig. 3.6), which are presumed not to play crucial roles in this protein's function.

Generally, the degree of change in a protein's primary structure is found to be inversely proportional to the biological importance of each residue. The most variable residues are those that occur on the surface of a protein but are not involved in functional interactions with other molecules (Chaps. 6 and 7). The most conserved amino acid residues are those that are most directly involved in the biological function of the protein, for example, the residues in cytochrome *c* that interact directly with the heme group (Figs. 3.6 and 3.8) and the active sites of enzymes. The same considerations apply to gene sequences. The untranslated regions of genes, particularly the introns, vary much more than the regions coding for proteins. Within the regions coding for protein, the most frequent nucleotide changes are those that do not alter the amino acid sequence.

The neutral mutation rate differs for each nucleotide in a gene and is usually a good indicator of the functional importance of each part of the amino acid sequence. Proinsulin is a good example (Fig. 2.14). The C peptide has evolved at a rate that is seven times more rapid than that of the A and B chains, which make up the functional hormone (Table 3.3). The C peptide is removed proteolytically from the middle of the proinsulin polypeptide chain after it has folded to its correct conformation. The primary role of the C peptide appears to be to ensure correct folding of the protein; it has no other known role, and other cross-links are able to function in the refolding of insulin in vitro. The greater rate of divergence of the C peptide than of the A and B chains, therefore, reflects the fewer constraints on its precise amino acid sequence, relative to the functional parts of the hormone.

All the preceding observations indicate that the type of changes at the molecular level that have occurred during evolution are those that are least likely to have functional consequences (and least likely to have been selected). Thus, the occurrence of primarily nonfunctional changes is most readily explained as being the result of the accumulation of neutral mutations. Natural selection at the molecular level seems primarily to be negative, weeding out the deleterious mutations that affect function.

Of course, functional changes have occurred during evolution, as evidenced by the diversity of organisms. This diversity is often not evident at the molecular level, in that proteins with the same function that are from different species usually have very similar properties. There are exceptions, however; the hemoglobins of vertebrates vary widely in the ways that their oxygen-binding properties are regulated (Sec. 8.4.3). For example, fish hemoglobins are used for respiration in the usual way, but they also secrete oxygen into the swim bladder and the eye in order to regulate buoyancy. This release of oxygen, which occurs in response to a decrease in the pH of the swim bladder, is known as the *Root effect* and does not occur in the hemoglobins of nonfish species. In another example, crocodiles are able to stay underwater for as long as an hour because their hemoglobins have evolved to liberate oxygen to cells only when absolutely required. Also, some birds are able to fly at a very high altitude because their hemoglobins have very high oxygen affinities. These are just a few of the ways that hemoglobins have evolved to permit species to occupy extreme environments, and such evolutionary changes would be expected to have been

hastened by natural selection. All of these functional differences can be attributed to mutational alterations of just a few residues (Sec. 8.4.3), however, and most of the evolutionary divergence that has occurred in the hemoglobins is believed to be neutral.

There are remarkably few other instances at the molecular level in which natural selection has had a positive effect in selecting *for* favorable mutations. One of the best candidates is the insulin of the guinea pig, which has evolved at a much greater rate than in other species. The guinea pig hormone has an unusually low biological potency but is present at relatively high levels. Positive selection may be enhancing a novel biological property of the insulin at the expense, perhaps, of its potency as an insulin. The two related hormones glucagon and pancreatic polypeptide have also evolved in the guinea pig at greater than normal rates, giving the guinea pig a number of biochemical peculiarities. Many of these apparent anomalies, however, can be explained by an alternative evolutionary origin for the guinea pig.

In another possible instance, two groups of mammals (ruminants and colobine monkeys) have independently evolved a fermentative foregut in which the enzyme lysozyme apparently digests bacteria. The lysozymes from these two groups share certain similarities in their functional properties and in their amino acid sequences that are not present in other lysozymes. These unusual lysozymes have evolved at twice the normal rate, suggesting that at least several of the amino acid changes are functional and were selected for.

The most dramatic evidence for positive selection for functional differences is found in the protein inhibitors of proteolytic enzymes. These proteinase inhibitors act by binding at the active site of the proteolytic enzyme and blocking its access to substrate (Sec. 9.3.2). The evolutionary variation observed in closely related proteinase inhibitors and their genes is just the opposite of that usually observed, in that the functionally important regions have changed the most. The most variable parts of the genes are those coding for the protein, in which most of the nucleotide replacements change the amino acid coded for. Those residues known to interact directly with the proteolytic enzymes have changed the most. Some of the inhibitors have been shown to be specific for different proteases. A corresponding hypervariability of the active site regions of certain proteolytic enzymes has also been observed, so some proteases and their inhibitors may be coevolving by positive selection.

It is probable that other examples of the role of positive selection pressure for functional differences will be discovered, but most evolutionary divergence of proteins is probably of the neutral variety and of no functional significance.

References

Evolutionary rate at the molecular level. M. Kimura. *Nature* 217:624–626 (1968).

Non-Darwinian evolution. J. L. King and T. H. Jukes. *Science* 164:788–798 (1969).

Biochemical peculiarities of the guinea pig and some examples of convergent evolution. J. C. Wriston. *J. Mol. Evol.* 17:1–9 (1981).

Guinea pig preproinsulin gene: an evolutionary compromise? S. J. Chan et al. *Proc. Natl. Acad. Sci. USA* 81:5046–5050 (1984).

Species adaptation in a protein molecule. M. F. Perutz. *Adv. Protein Chem.* 36:213–244 (1984).

Adaptive evolution in the stomach lysozymes of foregut fermentors. C. B. Stewart et al. *Nature* 330:401–404 (1987).

Ovomucoid third domains from 100 avian species: isolation, sequences, and hypervariability of enzyme-inhibitor contact residues. M. Laskowski, Jr., et al. *Biochemistry* 26:202–221 (1987).

Functional evolutionary divergence of proteolytic enzymes and their inhibitors. T. E. Creighton and N. J. Darby. *Trends Biochem. Sci.* 14:319–324 (1989).

Concerted evolution of ruminant stomach lysozymes. Characterization of lysozyme cDNA clones from sheep and deer. D. M. Irwin and A. C. Wilson. *J. Biol. Chem.* 265:4944–4952 (1990).

Is the guinea pig a rodent? A. D. Graur et al. *Nature* 351:649–652 (1991).

3.3.2 Variation within Species

It is usually possible to describe a human insulin, a bovine ribonuclease, or a horse cytochrome *c* because members of a species tend to have the same genes and proteins. The reason for this is genetic, due to the finite number of individuals in any species. As mentioned earlier, each gene in an individual has only a moderate probability of being passed on to the next generation. When the population is stable in size, this probability is 0.75 on average, in which case it is improbable that any particular copy of a gene will be passed on for very many generations, even if there are any typically moderate selective pressures. As a consequence, all copies of a particular gene that are present at any instant in a population are likely to have descended from a single copy that was present a limited number of generations previously; in this case, the genetic variation among the copies of a gene in the individuals of a population is limited to that which has arisen by mutation in the meantime.